Effect of Neuromodulation on Performance in Game Playing: A Modeling Study

Derrik E. Asher, Andrew Zaldivar, and Jeffrey L. Krichmar

Abstract— **Neuromodulators can have a strong effect on how organisms learn and compete for resources. Neuromodulators, such as dopamine (DA) and serotonin (5-HT), are known to be important in predicting rewards, costs, and punishments. To better understand the effect of neuromodulation on decision-making, a computational model of the dopaminergic and serotonergic systems was constructed and tested in games of conflict. This neural model was based on the assumptions that dopaminergic activity increases as expected reward increases, and serotonergic activity increases as the expected cost of an action increases. Specifically, the neural model guided the learning of an agent that played a series of Hawk-Dove games against an opponent. The model responded appropriately to changes in environmental conditions or to changes in its opponent's strategy. The neural agent became Dove-like in its behavior when its dopaminergic system was compromised, and became Hawk-like in its behavior when its serotonergic system was compromised. Our model suggests how neuromodulatory systems can shape decision-making and adaptive learning in competitive situations.**

Index Terms— **Dopamine, Serotonin, Game Theory, Computational Neuroscience, Decision-Making**

I. INTRODUCTION

Neuromodulators, such as dopamine (DA) and serotonin (5-HT), are known to be important in predicting rewards, costs, and punishments.

Dopamine activity (DA), which originates in the ventral tegmental area (VTA) and the substantia nigra (SN), appears to be linked to expected reward [1], and incentive salience or "wanting" [2]. Alternatively, it has been proposed that DA is involved with the discovery of new actions, and influenced action-outcome contingencies [3]. In all of these variants, DA is an important signal for the acquisition of salient, value-laden objects.

Serotonin (5-HT), which originates in the Raphe nucleus, appears to be related to cognitive control of stress, social interactions, and risk taking behavior [4, 5]. The structures that are innervated by 5-HT and their connecting circuits modulate the behavioral response to threats and risks, that is, behaviors that are typically thought to reflect the anxiety state

of the organism [4]. Whereas dopamine is tied to the expected reward of a given decision, serotonin could be thought of as related to the expected cost of a decision.

Game theory has been useful for understanding risk-taking and cooperation [6]. Of particular interest are studies in which neuromodulators were depleted or altered, while subjects play games. In one study, subjects, who were 5-HT depleted through dietary changes, cooperated less in a Prisoner's Dilemma game [7]. In an ultimatum game study, 5-HT depleted subjects tended to reject monetary offers more than control subjects when they deemed the offers to be unfair [5]. Manipulations of dopamine levels can significantly alter the ability to assess rewards in humans [8].

To better understand the roles of dopamine and serotonin during decision-making in games of conflict, we developed a computational model of neuromodulation and action-selection, based on the assumption that dopamine levels are related to the expected reward of an action, and serotonin levels are related to the expected cost of an action. An agent, whose behavior was guided by the neural model, played the Hawk-Dove game against different opponents. In the Hawk-Dove game, players must choose between confrontational and cooperative tactics [6, 9]. The results of these modeling experiments suggest a mechanism of how the neuromodulatory systems adapt behavior and decision-making under varying environmental conditions and opponents.

II. METHODS

A. Hawk-Dove Game

Two agents played a variant of the Hawk-Dove game [10]; one agent was a computer model, whose actions were guided by a rigid strategy (*Opponent*), the other agent was a neural network model that mimicked the effects of serotonin and dopamine on action selection and learning (*Neural*).

At the start of the game, both agents were randomly placed in an area where there was a territory of interest (TOI). Assuming both agents moved at the same speed, they would have an equally likely chance of reaching the TOI first. The agent that reached the TOI first could open with an *Escalate* (i.e. an aggressive, confrontational tactic) or a *Display* (i.e. a nonviolent, cooperative tactic). The second agent to reach the TOI could respond with an *Escalate* or *Display*. The payoff matrix for this game is given in Table I. If both agents *Escalate*, they received a penalty that was either a serious injury (large penalty) or just a scratch (small penalty). The probability of serious injury was set to 0.25 or 0.75 at the start

of the game. If both agents displayed, they shared the TOI resource. If one agent selected to Escalate and the other to Display, the agent that escalated got the entire resource.

	B. Escalate	B. Display
A. Escalate	A: (V–D)/2, B: (V–D)/2	A: V, B: 0
A. Display	A: 0, B: V	A: V/2, B: V/2

V is the value of the resource and is set to 0.60. D is the damage incurred when both players escalate. D is set to 1.60 for a serious injury and 0.62 for a scratch. The probability of a serious injury varies under different game conditions.

B. Opponent Agent

The *Opponent* followed one of 3 strategies. In one strategy, referred to as the *Statistical* model, the agent had a probability of escalation independent of the *Neural* agent's tactics, which was set at the beginning of the game to 0.25 or 0.75. In the second strategy, referred to as *Tit-For-Tat* (*TT*), the computer model always repeated the *Neural* agent's previous move. The only exception to this rule was if the *Opponent* agent reached the TOI first in the opening game, in which the *Opponent* opened with a Display. *TT* is a simple, yet effective strategy in game theory, which has shown to be successful in game playing tournaments [9]. The third strategy referred to as *Win-Stay, Lose-Shift (WSLS)*, the *Opponent* agent would win and stay with the same action in one of two possibilities; the *Opponent* agent's *Escalation* is met with the *Neural* agent's *Display* or the *Opponent* agent's Display is matched by a *Neural* agent's Display, otherwise the *Opponent* agent resorted to a lose and shift action [11]. As with the *TT* strategy, the *WSLS* opponent would open with a Display action if it arrived at the TOI first on the first game.
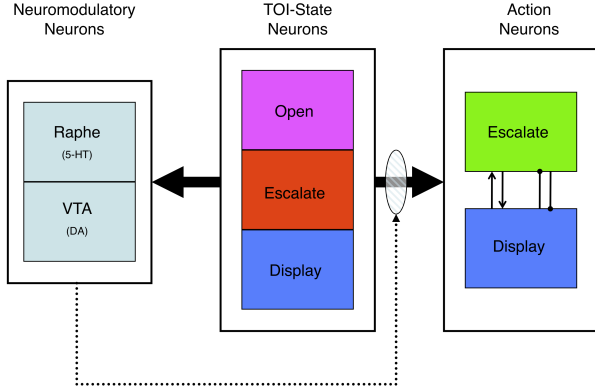

Fig. 1. Architecture of the neural model (two *Neuromodulatory*: Raphe and VTA; three *TOI-State*: Open, Escalate, and Display; and two *Action*: Escalate and Display). The solid arrows extending from the *TOI-State* neurons represent all-to-all connections. The shaded oval and dotted arrows represent plasticity of the pathways. Within the *Action* neurons, the line with the arrow at the end represent excitation, and the line with the dot at the end represent inhibition.

C. Neural Agent

A neural network controlled the behavior of the *Neural* agent. The neural network had three areas: *TOI-State*, *Action*, and *Neuromodulator* (Fig. 1).

There were three neurons in TOI-State area that corresponded to the possible tactics the *Opponent* agent could

take: 1) *Open*. The *Neural* agent reached the TOI first. 2) *Escalate*. The *Opponent* agent reached the TOI first and escalated a conflict. 3) *Display*. The *Opponent* agent reached the TOI first but did not start a conflict. The equation for the activity of each of these neurons (n_i) was set based on the current state of the TOI:

$$n_i = \begin{cases} 0.75 + rnd(0.0,\ 0.25); & i = TOIState \\ rnd(0.0,\ 0.25); & Otherwise \end{cases} \quad (1)$$

where $rnd(0.0, 0.25)$ was a random number uniformly distributed between 0.0 and 0.25.

There were two neurons in the action area: 1) *Escalate*. The *Neural* agent escalated a conflict. 2) *Display*. The *Neural* agent did not start a conflict or retreated if the *Opponent* agent escalated. The neural activity was based on synaptic input from *TOI-State* neurons and neuromodulation.

There were two neurons in the neuromodulator area: 1) *Raphe*. A simulated raphe nucleus and the source of serotonergic neuromodulation. 2) *VTA*. A simulated ventral tegmental area and the source of dopaminergic neuromodulation. The activity of these neurons was based on synaptic input from the *TOI-State*.

From	To	Initial Weight	Plastic	Phasic Neuromodulation
TOI-State	*Action*	0.1	Y	Y
TOI-State	*Neuromodulator*	0.1	Y	N
Action Escalate	*Action Display*	0.1	N	N
Action Escalate	*Action Display*	-0.1	N	Y
Action Display	*Action Escalate*	0.1	N	N
Action Display	*Action Escalate*	-0.1	N	Y

The synaptic connectivity of the network is shown in Fig. 1 and in Table II. The connections, given by each row in Table II, were all-to-all (i.e. every pre-synaptic neuron connected with every post-synaptic target). Some of these connections were subject to synaptic plasticity, and other connections were subject to phasic neuromodulation, where the activity of the *Neuromodulator* area could affect the synaptic efficacy. The neural activity was simulated by a mean firing rate neuron model, where the firing rate of each neuron ranged continuously from 0 (quiescent) to 1 (maximal firing). The equation for the mean firing rate neuron model was:

$$s_i(t) = \rho_i s_i(t-1) + (1 - \rho_i)\left(\frac{1}{1 + \exp(-5I_i(t))}\right) \quad (2)$$

where t was the current time step, s_i was the activation level of neuron i, ρ_i was a constant set to 0.1 and denoted the persistence of the neuron, and I_i was the synaptic input. The synaptic input of the neuron was based on pre-synaptic neural activity, the connection strength of the synapse, and the amount of neuromodulator activity:

$$I_i(t) = rnd(-0.5, 0.0) + \sum_j nm(t-1)w_{ij}(t-1)s_j(t-1) \quad (3)$$

where w_{ij} was the synaptic weight from neuron j to neuron i,

and *nm* was the level of neuromodulator at synapse *ij*. Phasic neuromodulation had a strong effect on action selection and learning. During phasic neuromodulation, synaptic projections from sensory systems (e.g. visual, auditory, etc) and inhibitory neurons were amplified relative to recurrent or associational information [12-14]. To simulate the effect of phasic neuromodulation, inhibitory and sensory connections were amplified by setting *nm* (equation 3) to ten times the combined average activity of the simulated Raphe, and VTA neurons. Otherwise, *nm* was set to one for recurrent or association connections. The last column of Table II lists connections amplified by phasic neuromodulation.

Action selection depended on the summed activity of the *Action* neurons after the neural agent reached the TOI. When the *Neural* agent reached the TOI, neural activities of the *Action* and *Neuromodulator* neurons were calculated for ten time-steps (equations 1-3). The Action neuron with the largest total activity during those ten time-steps dictated the action to be taken (e.g. if the total Display activity was greater then Escalate, the agent Displayed).

After both the *Opponent* and *Neural* agents chose a tactic, a learning rule, which depended on the current activity of the pre-synaptic neuron, the post-synaptic neuron, the overall activity of the neuromodulatory systems and the payoff from the game, was applied to the plastic connections (Table II).

$$\Delta w_{ij} = \delta * nm(t-1)s_j(t-1)\left(s_i(t-1)\right)*R \quad (4)$$

where s_j was the pre-synaptic neuron, s_i was the post-synaptic neuron, δ was a learning rate set to 0.1, *nm* was the average activity of all neuromodulatory neurons, and *R* was the level of reinforcement based on payoffs and costs (equation 5). The pre-synaptic neuron (s_j) in equation 4 was the most active *TOI-State* neuron. The post-synaptic neuron (s_i) was either the most active *Action* neuron, the *Raphe* neuron, or the *VTA* neuron. Weights were normalized by the square root of sum of squared weights.

The level of reinforcement (R, equation 4) was:

$$R = \begin{cases} (\text{Reward} - \text{VTA}) - (\text{Cost} - \text{Raphe}); & \text{TOI - State} \rightarrow \text{Action connection} \\ \text{Reward} - \text{VTA}; & \text{TOI - State} \rightarrow \text{VTA connection} \\ \text{Cost} - \text{Raphe}; & \text{TOI - State} \rightarrow \text{Raphe connection} \end{cases} \quad (5)$$

where the Reward was the *Neural* agent's payoff from Table I divided by the maximum possible reward. It was assumed that serotonergic plasticity was based on the expected cost of an action and dopaminergic plasticity was based on the expected reward of an action. If the Raphe or VTA accurately predicted the respective cost or payoff of an action, learning ceased. The *Neural* agent's cost was 1 if seriously injured, the ratio of scratch to serious injury (i.e. 0.3875, Table I) if scratched, or zero otherwise.

D. Game Playing

A *game* consisted of both agents (*Opponent* and *Neural*) taking a single action in response to a TOI (i.e. Escalate or Display). At the start of each game the agents were randomly placed in a square grid (not occupying the same area) and

were modeled to approach the neutral TOI at the same speed. The agent that arrived at the neutral TOI first had the opportunity to take either of the two possible actions (Escalate or Display), and the agent that arrived second responded with one of the two possible actions. After each game, the payoff was calculated and the plastic connections were updated.

A *series* consisted of 100 games with a given parameter set (e.g. *Control* agent against the *TT* opponent with serious injury set to 0.75). At the start of each series, the neural network was initialized and the *Neural* agent was considered "naïve", that is, the weights of the network were set to their initial values (Table II). For each parameter set, the two agents played 100 Hawk-Dove series with a different random number seed.

III. RESULTS

During the course of a series, the *Neural* agent learned to adopt different strategies depending on the chance of serious injury and its *Opponent's* strategy. To ensure that these strategies did not occur by chance, 100 randomly behaving agents played against all three *Opponents*. The random agents had lesions (i.e. activity set to zero) of both the simulated *VTA* and *Raphe*, which resulted in no learning occurring (equation 4). The 95th percentile of a confidence interval for the random agents was used as the cutoff for gauging non-random behavior. This cutoff corresponded to the probability of selecting a particular action in response to a given *TOI-State* greater than 65% or less than 35% of the time.
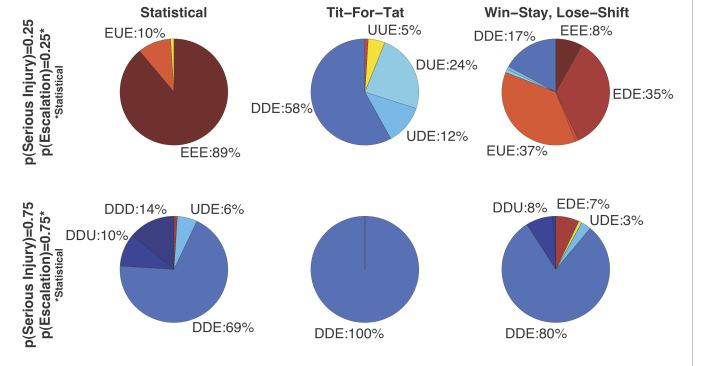

Fig. 2. The pie charts show the proportion of probable actions taken by the *Neural* agent in 100 series of games. There are three *TOI-States* (Open, Escalate, and Display), and three outcomes the *Neural* agent can commit to: Escalate (E), Display (D) or Undecided (U). Undecided represents random choice between 'E' and 'D'. The labels represent the *Neural* agent's response to the three *TOI-State* areas. Strategies that are *Dove-like* are displayed in blue, *Hawk-like* are displayed in red, and arbitrary strategies displayed in yellow.

A. Response to Environmental Change and Opponent Tactics

The *Control* agent adapted its behavior depending on its opponent's strategy and environmental conditions (Fig. 2). In response to a given TOI-state, the agent could respond randomly (i.e. within the 95% confidence), or significantly tend toward escalation or displaying. Therefore, there are a total of 27 possible outcomes the *Neural* agent can take with respect to the three different *TOI-States*. Only a few of these outcomes emerged in the simulations, and these outcomes are

represented in Fig. 2 as a triplet pairing (i.e., EEE, DDE, UDE, etc.). The first value in the triplet pairing corresponds to the expected action when the *TOI-State* is *Open*. The second represents the anticipated action when the *TOI-State* is *Escalate*. Lastly, the third value denotes the expected outcome when the *TOI-State* is *Display*. For instance, if the triplet pairing reads DDE, then that means the *Neural* agent tends to display when the TOI is *Open*, display when the *Opponent* escalated, and escalate when the *Opponent* displayed. These triple pairings are associated with a color spectrum, where aggressive outcomes (instances of 'E' in the triplet) are denoted red, passive outcomes (instances of 'D' in the triplet) are denoted in blue, and values that do not fall within either outcome (instances of 'U' in the triplet) are denoted in yellow. Against all three Opponent types, the Neural agent adopted *Hawk-like* behavior in "safe" environments, where the probability of a serious injury is 0.25 (top row, Fig. 2), and *Dove-Like* behavior in "harsh" environments, where the probability of a serious injury is 0.75 (bottom row, Fig. 2).

B. *Effect of Lesions on Behavior*

A set of simulated lesion experiments to the Raphe and VTA areas were carried out to assess whether an intact neuromodulatory system was necessary for the *Neural* agent to respond appropriately to changes in the environment and an opponent's strategy.

TABLE III
PERCENTAGE OF ESCALATION FOR THE NEURAL AGENT

	Control Agent		Raphe Lesion		VTA Lesion	
	Safe	*Harsh*	*Safe*	*Harsh*	*Safe*	*Harsh*
Statistical	97.65%	10.00%	99.06%	92.86%	34.79%	7.14%
TT	34.15%	13.64%	81.82%	81.82%	24.74%	12.50%
WSLS	93.22%	9.09%	96.88%	96.88%	20.93%	8.22%

A simulated lesion to the serotonergic system (*Raphe*) resulted in an inability to assess the cost of a decision. Table III depicts the total percentage of escalations across all games and series. Table III shows that the *Raphe* agent had a significantly higher percentage of escalation than other agents in the harsh environment per opponent (p << 0.0001, Wilcoxon Rank Sum). This suggests that when the Raphe area was removed, the agent behaved exclusively *Hawk-like*, even against the *TT* and *WSLS* opponents. By comparison, in Table III, the *Control* agent with the Raphe area intact—dropped its percentage of escalation as the *Opponent* became more sophisticated and the environment became harsher (p << 0.0001, Wilcoxon Rank Sum). Because the *Raphe* agent could not judge the cost of escalating, it relied solely on its ability to assess reward through its intact VTA area, resulting in Escalating where there is an expectation of larger rewards.

A simulated lesion to the dopaminergic system (*VTA*) revealed an inability to assess the reward of a decision. Table III shows that the *VTA* agent had a significantly lower percentage of escalation (i.e., higher percentage of displaying), than the other agents in the safe environment (p << 0.0001, Wilcoxon Rank Sum). The *VTA* agent behaved more *Dove-like* than the *Control* agent. In contrast, the *Control* agent, which had the VTA area intact, increased its percentage of

escalations as the environment became more forgiving (see Table III). These results imply that the *VTA* agent could assess the cost of escalating, but not the value of reward, and consequently adopted *Dove-like* behavior.

C. *Effect of Lesions on Payoffs*

The inability to assess cost due to a Raphe lesion not only impacted the *Neural* agent's ability to obtain optimal payoff for itself, but for its opponent as well (see Fig. 3). All *Opponents* against the *Raphe* agent received a lower payoff compared to the *Opponents* playing against any other *Neural* agent (p << 0.0001, Wilcox Rank Sum). For example, the *Control* agent and its *Opponent* received a higher payoff than the *Raphe* and its *Opponents* in all conditions (compare 3A with 3B).
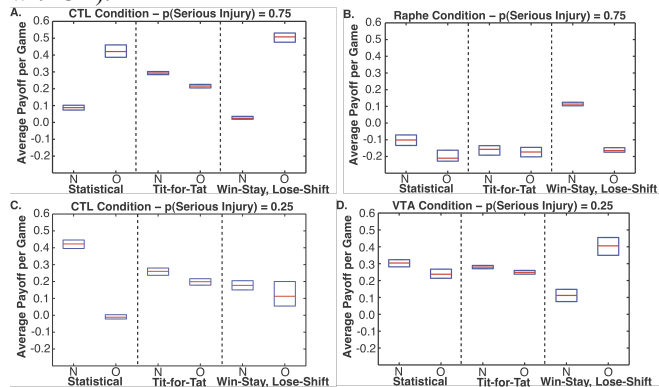

Fig. 3. Average payoff score for the *Neural* (N) and *Opponent* (O) agents. The *Neural* agent conditions include: A. Intact simulated neural network with learning capabilities (CTL). B. Lesion to the simulated Raphe nucleus (Raphe). C. Lesion to the simulated Ventral Tegmental Area (VTA). D. Lesion to both the Raphe and VTA (Dual). The columns correspond to the different *Opponent* strategies: Statistical, *Tit-For-Tat* (TT), and *WSLS* (WSLS). The red line in the box plots shows the median values and the blue ends to the box show the 25th and 75th percentiles.

Lesioning the VTA area resulted in lower payoffs for the *Neural* agent and higher payoffs for its opponent compared to that of controls (compare Fig. 3C with Fig. 3D). In safe environments, the *VTA* agent received significantly lower payoffs against the *Statistical* and *WSLS* opponents than the intact model (p << 0.0001, see Neural payoff in Statistical and WSLS in Fig. 3D). Because the *VTA* agent could not assess reward but could still assess cost, it adapted its behavior to a low-risk Dove strategy. This allowed its opponent to escalate without being penalized and receive maximum rewards at the *VTA* agent's expense, while the *VTA* agent received minimal injuries.

Random behavior resulted from the non-learning, dual lesion case. By definition, the *Dual* agent escalated spontaneously ~50% of the games in a series. This resulted in sub-optimal payoffs to both agents (p << 0.0001). *Dual* received a higher payoff than *Control* against *WSLS* opponent because *Hawk-like* behavior resulted in greater payoff to the *Neural* agent (p << 0.001, compare Fig. 3B: (WSLS) Neural to 3A: (WSLS) Neural and 3B: (WSLS) Neural), and random behavior is ~50% *Hawk-like*. Thus, *Dual* received a higher payoff than *Control* against *WSLS*, but this was an artifact of random actions against the WSLS strategy. The fact that *Dual*

utilized a fixed strategy illuminates the learning that took place as a result of the intact neuromodulatory system in *Control* (p << 0.001).

IV. DISCUSSION

In the present paper, we showed that an agent, whose behavior was guided by a computational model of neuromodulatory action, learned to adjust its strategy appropriately depending on environmental conditions or its opponent's strategy in the Hawk-Dove game. Lesions of the simulated neuromodulatory system resulted in perseverating behavior and detrimental performance to both the *Neural* agent and its *Opponent*. The model makes several predictions on how the action of neuromodulatory systems can lead to appropriate action selection in competitive and cooperative environments.

A. Neuromodulators Track Expected Rewards and Costs

In constructing the model, it was assumed that dopaminergic activity increased as expected reward increased, and that the serotonergic system increased as the expected cost of an action increased. When the Neural agent's dopaminergic system was compromised, it was unable to predict the payoff associated with an action and resorted to *Dove-like* behavior (VTA in Table III). In contrast, when the Neural agent's serotonergic system was compromised, it became *Hawk-like* in its behavior (Raphe, Table III). Indeed, the activity of neurons in the simulated neuromodulatory area reflected the differing rewards and costs associated with the different strategies adopted by the *Neural* agent.

Dopamine appears to be important for reward anticipation [15], and the "wanting" of things, that is, the motivation process in acquiring an object [2]. Therefore, having the dopamine activity related to the reward payoff in the game (equation 5) appeared to be a reasonable assumption.

On the other hand, serotonergic activity appears to modulate behavioral response to risks, stress, and threats [4, 16], as well as playing an important role in social anxiety in primates [17]. All of these risks and threats have a cost associated with them. Moreover, reduced serotonin transmission is associated with a release of aversive or punishing responses [18]. Therefore, it seemed reasonable to assume that serotonin activity is related to the expected cost of a given action.

Given these assumptions, the *Control* agent, adjusted its strategy appropriately depending on environmental conditions and on its *Opponent's* strategy (Fig. 2). For example, in situations where it was more likely to sustain a serious injury during a confrontation, the *Control* agent's behavior became more *Dove-like*. This resulted in an increase in Display and decrease in Escalate actions when playing against the *Statistical* or *WSLS* models (see Table III), as well as a decrease in random behavior when playing against a *TT* opponent (Fig. 2: middle column). The *Control* agent learned that there was an increased cost and decreased reward to be expected by escalating a confrontation when the probability of serious injury increased. No matter which *Opponent* the *Control* agent faced, it learned to alter its strategy to take advantage of a no cost escalation in response to its *Opponent* Display 1st action. This can be seen in Fig. 2 for all tactics that end in E (e.g. DDE or EDE).

The adaptive behavior demonstrated by the *Control* agent required an intact neuromodulatory system in which the agent could evaluate the expected cost and the expected reward of a given action. When the simulated dopaminergic or the simulated serotonergic systems were lesioned, the *Neural* agent's behavior became either completely *Hawk-like* (Raphe lesion), or completely *Dove-like* (VTA lesion). It is clear that extreme *Hawk-like* tactics can be viewed as uncooperative and is in agreement with behavioral studies in which serotonin levels were lowered [4, 19-21]. But it is less obvious how the VTA lesion is altering the agent's behavior.

In the present simulation study, lowering dopamine results in the *Neural* agent avoiding risks that lead to a higher payoff. These results are in agreement with a study in which a blockade of dopamine resulted in rats not making an extra effort of climbing over a barricade to get a high reward [22]. Moreover, a recent study has shown that individuals with a polymorphism that lowers levels of dopamine in the prefrontal cortex tended to take less risks in a gambling task [23]. Our results predict that by lowering dopamine levels, the agent loses its ability to assess forthcoming rewards, which results in risk averse behavior.

Our results predict that by lowering serotonin levels, the agent loses its ability to assess cost, and therefore achieving higher payoff is driving its actions. This results in *Hawk-like* behavior and is in agreement with recent human studies under similar conditions. In one study, humans played the Prisoner's Dilemma game (essentially equivalent to the Hawk-Dove game) under conditions that varied their tryptophan levels by Acute Tryptophan Depletion (ATD) [7]. Subjects defected against their opponent significantly more when their serotonin levels were low. This is the equivalent of escalating in the present Hawk-Dove game. In another ATD study where subjects played the Ultimatum Game, participants with depleted 5-HT levels rejected significantly more of the unfair offers [5].

B. Comparison to Other Models

Other computational models such as Evolutionary Algorithms and Reinforcement Learning have been effective in developing optimal strategies in games of conflict [24-27]. Soltani, Lee, and Wang showed that monkeys tend towards a WSLS strategy when the model utilized a Reinforcement Learning algorithm, which used the sequence of the monkey's choices for the given day [28]. Although our model shows many of these interesting adaptive behaviors, the main purpose of our approach was to better understand the role of neuromodulation in cooperative behavior. Game theory is one such method for studying this behavior. It may be of interest in the future to pit our neurobiologically inspired model against some of these reinforcement learning and evolutionary algorithms.

Computational models have been developed to emulate neuromodulatory processes during decision-making or action

selection [28]-[31]. One model integrated four neuromodulator systems into the temporal difference (TD) equation; with different parameters representing different neuromodulators; DA for reward prediction, 5-HT for discounting, norepinephrine (NE) for exploration/exploitation, and acetylcholine (ACh) for learning rate [29]. Daw, Kakade, and Dayan proposed a model in which dopamine and serotonin levels track predicted rewards and punishments [30]. This differs from our model in that punishments and rewards are not necessarily mutually inhibitory. Our model takes into consideration that an action could have independent costs and rewards associated with it (i.e., an action may have a high predicted reward, and a high predicted cost).

Although our model of neuromodulation has many similarities to these models, we designed our model with a specific hypothesis on the role of phasic neuromodulation. Specifically, (i) the common effect of the neuromodulatory systems is to drive an organism to be decisive when environmental conditions call for such actions, (ii) the main difference between neuromodulatory systems is the environmental stimuli that activate them, (iii) phasic neuromodulation can increase the signal to noise ratio of downstream neuronal targets by amplifying connections carrying sensory information and, (iv) phasic neuromodulation gates in learning [14, 31]. By using game theory to test our hypothesis, the present study extends our previous work by investigating the role of neuromodulation in learning.

Our model suggests a simple mechanism for adaptive behavior in competitive and cooperative situations. It is based on the assumptions that activity in the dopaminergic system is related to the expected reward of an action, and activity in the serotonergic system is related to the expected cost of an action.

The model makes the following predictions: 1) An intact neuromodulatory system is necessary for appropriate decision making and adapting to the environment in situations where cooperation is important. 2) Impairment to either the dopaminergic or serotonergic system will lead to perseverant behavior. Impairment of the dopaminergic system results in risk-averse behavior, and impairment of the serotonergic system results in risk-taking behavior. 3) Although dopamine and serotonin activity appears to be related to different expectations (e.g. predictive reward, anticipated cost), the action of these neuromodulators on downstream targets is similar, that is it governs decision-making.

REFERENCES

[1] W. Schultz, "Dopamine neurons and their role in reward mechanisms," *Curr Opin Neurobiol,* vol. 7, pp. 191-7, Apr 1997.

[2] K. C. Berridge, "Motivation concepts in behavioral neuroscience," *Physiol Behav,* vol. 81, pp. 179-209, Apr 2004.

[3] P. Redgrave and K. Gurney, "The short-latency dopamine signal: a role in discovering novel actions?," *Nat Rev Neurosci,* vol. 7, pp. 967-75, Dec 2006.

[4] M. J. Millan, "The neurobiology and control of anxious states," *Prog Neurobiol,* vol. 70, pp. 83-244, Jun 2003.

[5] M. J. Crockett*, et al.*, "Serotonin modulates behavioral reactions to unfairness," *Science,* vol. 320, p. 1739, Jun 27 2008.

[6] J. Maynard Smith and G. A. Parker, "The Logic of Animal Conflict," *Nature,* vol. 246, pp. 15-18, 1973.

[7] R. M. Wood*, et al.*, "Effects of tryptophan depletion on the performance of an iterated Prisoner's Dilemma game in healthy adults," *Neuropsychopharmacology,* vol. 31, pp. 1075-84, May 2006.

[8] B. Pleger*, et al.*, "Influence of dopaminergically mediated reward on somatosensory decision-making," *PLoS Biol,* vol. 7, p. e1000164, Jul 2009.

[9] R. Axelrod and W. D. Hamilton, "The evolution of cooperation," *Science,* vol. 211, pp. 1390-6, Mar 27 1981.

[10] J. Maynard Smith, *Evolution and the theory of games*: Cambridge University Press, 1982.

[11] M. Nowak and K. Sigmund, "A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game," *Nature,* vol. 364, pp. 56-8, Jul 1 1993.

[12] M. E. Hasselmo and J. McGaughy, "High acetylcholine levels set circuit dynamics for attention and encoding and low acetylcholine levels set dynamics for consolidation," *Prog Brain Res,* vol. 145, pp. 207-31, 2004.

[13] M. Kobayashi*, et al.*, "Selective suppression of horizontal propagation in rat visual cortex by norepinephrine," *Eur J Neurosci,* vol. 12, pp. 264-72, Jan 2000.

[14] J. L. Krichmar, "The Neuromodulatory System – A Framework for Survival and Adaptive Behavior in a Challenging World," *Adaptive Behavior,* vol. 16, pp. 385-399, 2008.

[15] W. Schultz*, et al.*, "A neural substrate of prediction and reward," *Science,* vol. 275, pp. 1593-9, Mar 14 1997.

[16] J. Amat*, et al.*, "Previous experience with behavioral control over stress blocks the behavioral and dorsal raphe nucleus activating effects of later uncontrollable stress: role of the ventral medial prefrontal cortex," *J Neurosci,* vol. 26, pp. 13264-72, Dec 20 2006.

[17] K. K. Watson*, et al.*, "Serotonin transporter genotype modulates social reward and punishment in rhesus macaques," *PLoS ONE,* vol. 4, p. e4156, 2009.

[18] R. Cools*, et al.*, "Serotoninergic regulation of emotional and behavioural control processes," *Trends Cogn Sci,* vol. 12, pp. 31-40, Jan 2008.

[19] N. Schweighofer*, et al.*, "Low-serotonin levels increase delayed reward discounting in humans," *J Neurosci,* vol. 28, pp. 4528-32, Apr 23 2008.

[20] S. C. Tanaka*, et al.*, "Serotonin differentially regulates short- and long-term prediction of rewards in the ventral and dorsal striatum," *PLoS ONE,* vol. 2, p. e1333, 2007.

[21] S. D. Hood*, et al.*, "Acute tryptophan depletion. Part I: rationale and methodology," *Aust N Z J Psychiatry,* vol. 39, pp. 558-64, Jul 2005.

[22] F. Denk*, et al.*, "Differential involvement of serotonin and dopamine systems in cost-benefit decisions about delay or effort," *Psychopharmacology (Berl),* vol. 179, pp. 587-96, May 2005.

[23] P. Roussos*, et al.*, "Planning, decision-making and the COMT rs4818 polymorphism in healthy males," *Neuropsychologia,* vol. 46, pp. 757-63, Jan 31 2008.

[24] T. W. Sandholm and R. H. Crites, "Multiagent reinforcement learning in the Iterated Prisoner's Dilemma," *Biosystems,* vol. 37, pp. 147-66, 1996.

[25] V. P. Crawford, "Adaptive Dynamics in Coordination Games," *Econometrica,* vol. 63, pp. 103-143, 1995.

[26] S. Chong and X. Yao, "Behavioral diversity, choices and noise in the iterated prisoner's dilemma," *IEEE Transactions on Evolutionary Computation,* vol. 9, pp. 540-551, 2005.

[27] P. G. Harrald and D. B. Fogel, "Evolving continuous behaviors in the Iterated Prisoner's Dilemma," *Biosystems,* vol. 37, pp. 135-45, 1996.

[28] A. Soltani*, et al.*, "Neural mechanism for stochastic behaviour during a competitive game," *Neural Netw,* vol. 19, pp. 1075-90, Oct 2006.

[29] K. Doya, "Metalearning and neuromodulation," *Neural Netw,* vol. 15, pp. 495-506, Jun-Jul 2002.

[30] N. D. Daw*, et al.*, "Opponent interactions between serotonin and dopamine," *Neural Netw.,* vol. 15, pp. 603-616, 2002.

[31] B. R. Cox and J. L. Krichmar, "Neuromodulation as a Robot Controller: A Brain Inspired Design Strategy for Controlling Autonomous Robots," *IEEE Robotics & Automation Magazine,* vol. 16, pp. 72-80, 2009.